THE U.S. GROUP ON EARTH OBSERVATIONS

DRAFT COMMON FRAMEWORK FOR EARTH-OBSERVATION DATA

Table of Contents

Background

Purpose of the Common Framework

Each year, Federal agencies invest billions of dollars in civil Earth observations. Through these investments, the U.S. Government ensures that the Nation's decision makers, emergency responders, scientists, businesses, farmers, and a wide array of other stakeholders have the information they need about climate and weather, disaster events, land-use change, ecosystem health, natural resources, and many other characteristics of the Earth system. Taken together,

Earth observations provide the indispensable foundation for advancing and sustaining the economic, environmental, and social well-being of the United States.

While Earth-observation systems typically collect data for a specific purpose, these data are often also useful in applications unforeseen during development of the systems. It is essential to manage and preserve these data so that users can find, evaluate, understand, and utilize the data in new and unanticipated ways. The wide range of scientific and observation efforts across the Federal agencies, and the diversity of data types collected as a result, require a data-management approach that can be both applied broadly across these data and tailored to particular needs.

The U.S. Government is committed to making data from civil Earth-observation assets freely available to all users. Even when there are no financial barriers, however, technical impediments to utilizing these data to their full potential often exist. In order to minimize these impediments, the *Common Framework for Earth-Observation Data* (CFEOD, also referred to in this document as "*the Common Framework"*) provides Federal agencies with a recommended set of standards and practices to follow. Federal agencies can follow these best practices as they develop new observing systems or modernize their existing collections of data in order to enhance the discoverability, accessibility, and usability of Earth-observation data. By standardizing the protocols for finding, accessing, and using Earth-observation data, the Common Framework will make it easier to obtain and assemble data from diverse sources for improved analysis, understanding, decision-making, community resilience, and commercial uses.[1]

[1] To use a plumbing analogy, the recommendations in this document address the pipes and pumps used to deliver water from its source. The purpose of the CFEOD is not to build the faucets and spigots actually used to access the water, but rather to standardize parameters such as pipe diameter. This approach makes it so that people designing and building faucets and spigots know how to connect to the plumbing and makes it easier to design systems that combine water from multiple sources since the pipes are compatible, encouraging creativity and new innovations.

The standards recommended in this document are not new. In fact, they are all existing standards that have been endorsed by the Federal Geographic Data Committee (FGDC)[2], an interagency committee within the U.S. Government that promotes the coordinated development, use, sharing, and dissemination of geospatial data on a national basis and are also found in the Global Earth Observation System of Systems (GEOSS) Standards and Interoperability Registry.[3] What is new about the Common Framework is the recommendation to use a focused subset of those standards within that list in order to increase data interoperability among Federal agencies. In addition, the Common Framework provides extra detail in pointing to implementations of those standards.

Target User of the Common Framework

As described above, this document is intended to provide guidance to data providers for improving and standardizing data-management practices in Federal agencies, and communicating to the users accessing that data what standards will be used. Following the recommendations by adopting these standards will not alone serve to deliver information derived from Earth observations to decision makers. The primary intent of the Common Framework is to reach those handling data within and outside of Federal agencies, in order to help these researchers and practitioners use the data for research purposes and/or create value-added products such as data portals, visualizations, and other tools. While this document does not provide guidance to all end-users of data, the expectation is that they will benefit by easing the creation and enhancing the utility of the tools that they use to interact with Earth-observation data.

[2] The full list of FGDC Endorsed Standards can be found at: http://www.fgdc.gov/standards/fgdc-endorsed-external-standards
[3] The GEOSS Standards and Interoperability Registry can be found at: http://seabass.ieee.org/groups/geoss/index.php

Scope of the Common Framework

The recommendations in this document are limited to the management of Earth-observation data. The Common Framework adopts the definition of Earth-observation data used by the 2013 *National Strategy for Civil Earth Observations*:

> "[We] use the terms "data" and "Earth observations" interchangeably to mean geo-referenced digital information about Earth, including the observations, metadata, imagery, derived products, data-processing algorithms (including computer source code and its documentation), and forecasts and analyses produced by computer models. Non-digital data, published papers, preserved geological or biological samples, or other media that have not been digitized are not included in this definition…"

This document originated as a mechanism to coordinate data-management improvements carried out through the Big Earth Data Initiative (BEDI, described in Appendix 1). This document was originally intended to give clear guidance to agencies on which practices to use when managing data under the auspices of BEDI. However, because the recommendations had considerable interagency support and may be applicable to data management at Federal agencies more broadly, the Office of Science and Technology Policy (OSTP) decided to release the Common Framework publicly for input. The Common Framework does not call for mandating the use of its endorsed practices, but can serve as a reference for data managers and users on a suggested subset of the many available standards with which they may want to familiarize themselves. The idea of the Common Framework is to narrow the scope of what is permissible for BEDI, and to recommend that others adopt the same standards in order to maximize interoperability. The recommendations in this document are intended to be broadly applicable, but not comprehensive.

The standards and protocols herein are endorsed for use on a wide range of datasets, but there are types of observations and special cases for which other practices will be necessary.

Structure of the Common Framework

The aspects of data management considered in this document are:

- Data Search and Discovery Services
- Data-Access Services
- Data Documentation
- Common Formats and Vocabularies

Within the sections about each aspect, the recommended best practices are discussed at three levels:

- *Standards and Protocols* - officially endorsed standards for use in Earth-observation data management.
- *Methods and Practices* - recommended ways to use the endorsed standards, including making data open by default per Project Open Data guidelines (EO M-13-13).
- *Implementations* - available software to use in realizing the standards and examples of the use of the standards at Federal agencies. These are not endorsements but are pointers for users of this document to find more information.

Data Search and Discovery Services

Introduction

Searching for relevant data is often the first step in addressing a new problem or research question. Federal agencies should seek to ensure that Earth observations are readily discoverable by the diverse community of domestic and international information providers and users. Data collected for a particular purpose can only be reused for novel purposes if those unanticipated users are able to know the data exists and find it. This section provides guidance on the creation and organization of searchable catalogs of data, which are created by harvesting metadata formatted in a standardized structure.

Typically, an organization will establish some sort of web page or portal with a human-oriented user interface to allow individuals to find data. This is a useful first step, but it is by no means sufficient – modern catalogs also need to be searchable by automated software including search engines, other catalogs, and decision-support software. These catalogs need to provide the ability to run various kinds of searches and get metadata records or summarization of metadata in the results. Accordingly, the *National Strategy for Civil Earth Observations* called for the establishment of formal standards-based catalog services which would enable commercial search engines to index data holdings.

The standards below, properly applied, can be used to meet the following Data Management Principles from the international Group on Earth Observations Strategic Plan (see Appendix 1):

- **DMP-1**: Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions, including licenses, will be clearly indicated.
- **DMP-10**: Data will be assigned appropriate persistent, resolvable identifiers to enable documents to cite the data on which they are based, and to enable data providers to receive acknowledgement of use of their data.

Standards and Protocols

For many users, the search for information begins with one of the commercial search engines that crawl the entire indexed Web. Search engines typically work by indexing the text on a page, and so cannot detect that a page is about Earth-observation data without appropriate text identifiers. Therefore, the Common Framework recommends:

- Including schema.org (https://schema.org) tags in dataset landing pages. Schema.org creates, maintains, and promotes schemas for structured data on the Internet. These tags can convey information such as:
    - geographic coverage: https://schema.org/GeoCoordinates
    - temporal coverage: https://schema.org/datasetTimeInterval
    - other dataset attributes: https://schema.org/Dataset

While many users will use external search engines for search and discovery, many others will begin their search in published catalogs. Thus, as well as opening data records to external search engines, data providers should create their own searchable catalogs.

The Common Framework recommends one or both of the following Application Programming Interfaces (APIs) to allow computer programs to search catalogs:

- Open Geospatial Consortium (OGC, http://www.opengeospatial.org/) Catalog Service for the Web (CSW, http://www.ogcnetwork.net/node/630) - OGC Catalogue interface standards specify the interfaces, bindings, and framework for defining application profiles required to publish and access digital catalogues of metadata for geospatial data, services, and related resource information.
    - CSW is specifically designed for searching geographic data.
    - CSW allows searches to specify locations and time periods of interest.
    - CSW supports the metadata standard recommended in the Data Documentation section.
- OpenSearch (http://www.opensearch.org/) - Description: OpenSearch is a collection of simple formats for the sharing of search results.
    - OpenSearch is generically applicable to any type of data.
    - OpenSearch extensions to support searching by location or time are currently in draft form.

To allow the combination of project- or organization-specific catalogs into more-general catalogs by harvesting metadata, the Common Framework recommends:

- Open Archives Initiative Protocol for Metadata Harvesting (OAI/PMH, https://www.openarchives.org/pmh/) - Description: OAI-PMH is a low-barrier mechanism for repository interoperability.
    - Allows a catalog to harvest all or part of the metadata from another catalog.

    - Harvest can request only recent changes, or only particular collections.

To enable specific datasets to be cited in journal articles, documents, web pages, or workflow descriptions, The Common Framework recommends:

- Digital Object Identifiers (DOIs, https://www.datacite.org/) assigned to datasets

  – Description: DataCite assigns persistent identifiers to datasets to increase data accessibility.

    - A DOI is equivalent to a serial number for a dataset.
    - The assigned DOI stays the same even if data are moved to another web site or organization.
    - DOIs should be assigned at an fairly coarse level of granularity (for example, one DOI for an entire time series of data, rather than separate DOIs for every single measurement).

Methods and Practices

Catalog registration method:

- Provide URL of Web Accessible Folder (WAF) containing International Organization for Standardization (ISO) metadata as described in the Data Documentation section.
    - Legacy data using FGDC Content Standard for Digital Geospatial Metadata (CSDGM) is accepted for legacy datasets that cannot comply with newly defined and adopted standards.

Persistent Identifier Use:

- When possible, values for metadata fields should also include persistent, resolvable identifiers. For example, when a person's name is given, if that person has an existing persistent identifier, then that identifier along with the establishing authority should also be provided.

Implementations

Software

- CKAN (open source) - http://ckan.org/
- Geoportal Server (open source) - https://github.com/Esri/geoportal-server/
- Geonetwork (open source) - http://geonetwork-opensource.org/
- GeoTools (open source toolkit that implements OGC-CSW and other OGC protocols) - http://www.geotools.org/

Instances

- National Oceanic and Atmospheric Administration (NOAA) Data Catalog (open source)- https://data.noaa.gov/, CKAN
- National Aeronautics and Space Administration (NASA) ECHO (open source) - https://earthdata.nasa.gov/echo
- United States Geological Survey (USGS) ScienceBase, GeoTools (open source) - http://www.sciencebase.gov/

Testing

- Unit tests (*e.g.*, catalog responds with correct number of records for query; info from each record is correct).
- Interoperability tests (*e.g.*, catalog client is successfully able to communicate with catalog server; metadata harvest or distributed search works correctly).
- Scenario-based tests (*e.g.*, ability to find relevant data to address particular scenario).
    - Example: U.S. Integrated Ocean Observing System (IOOS®) scenario tests - github.com/ioos/system-test

Metrics

Note: Metrics will be measured by automatically gathering statistical information from Agency data catalogs.

- Number of datasets listed in each catalog
- Number of datasets that meet the U.S. Chief Information Officer qualifications for Open Data: publicly accessible under an open license with no restrictions on reuse. See the full Open Data Principles at https://project-open-data.cio.gov/principles
- Percentage of Earth Observations Assessment (EOA) Observing Systems listed in one of the Agency data catalogs
- Percentage of datasets in catalog with one or more of the Common Framework Data Access methods offered

Data-Access Services

Introduction

Data-access services focus on the ways that users retrieve data for exploration, analysis, and decision-making. In the past, users might have been limited to downloading an entire data set, or data providers might have required the use of project-specific portals or web applications as the sole method for data access. The Common Framework aims to encourage data providers to offer services and application programming interfaces (APIs) that provide more-advanced access to data sets, including subsetting, aggregation, visualization, or other methods. These methods return data more relevant to the user's specific problem. An additional benefit is that as more data providers offer the same standards-based services, it becomes easier to obtain and combine data from multiple sources without additional programming or post-processing.

The standards below support the following Data Management Principle from the GEO Strategic Plan:

- **DMP-2**: Data will be accessible via online services, including, at minimum, direct download, but preferably user-customizable services for visualization and computation.

Standards and Protocols

Many Earth observations are “georeferenced,” meaning that the location on Earth of the observation is described by Earth-based coordinates like latitude and longitude, as opposed to a street address or political boundaries. “Raster images” are images that are structured as pixel-by-pixel data, as opposed to a set of line segments and shapes. The recommended data-access

services for map-based visualizations of georeferenced data, or direct access to raster imagery are:

- OGC Web Map Service (WMS) – Description: the OGC WMS provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases.
    - http://www.opengeospatial.org/standards/wms
- OGC Web Map Tile Service (WMTS) – Description: the OGC WMTS protocol serves map tiles of spatially referenced data using tile images with predefined content, extent, and resolution.
    - http://www.opengeospatial.org/standards/wmts

When data is accessed in numerical rather than image form, it is often structured into a grid by location, especially for remote sensing data. For regularly gridded data, where the elements within the grid are the same size, the recommendation is:

- Data Access Protocol (DAP) – Description: OPeNDAP is a data-transmission protocol designed specifically for science data.
    - http://www.opendap.org/support
    - https://earthdata.nasa.gov/standards/data-access-protocol-2
- OGC Web Coverage Service (WCS) – Description: OGC WCS offers multi-dimensional coverage data for access over the Internet.
    - http://www.opengeospatial.org/standards/wcs

Sometimes the data is in an unstructured grid, meaning that each element cannot be described by a simple set of coordinates because they are not of uniform size. For data in such a form, the recommendation is:

- *UGRID* – Description: metadata conventions for scientific data using unstructured grids.
    - https://github.com/ugrid-conventions/ugrid-conventions/blob/v0.9.0/ugrid-conventions.md

For *in situ* data, meaning data taken at the location of observation by one or more stationary or moving sensors:

- OGC Sensor Observation Service (SOS) – Description: the SOS standard defines a Web-service interface that allows querying observations, sensor metadata, and representations of observed features for managing data in an interoperable way.
    - http://www.opengeospatial.org/standards/sos
- OGC Web Feature Service (WFS) – Description: WFS allows users to retrieve or modify specific data relevant to the user.
    - http://www.opengeospatial.org/standards/wfs
- DAP with discrete sampling geometries.
    - https://www.nodc.noaa.gov/data/formats/netcdf/

For data that is not geo-referenced, but is described by geographic features such as political boundaries, roads, *etc.*:

- OGC Web Feature Service (as noted above).

For tabular data, meaning data structured in tables (*e.g.*, where each column is a different piece of information) rather than grids:

- TableDAP – Description: TableDAP allows the use of the OPeNDAP constraint/selection protocol to request data subsets, graphs, and maps from tabular datasets.
    - http://coastwatch.pfeg.noaa.gov/erddap/tabledap/

Methods and Practices

The following are additional suggested methods or practices to supplement the recommendations above.

- Adopt the WMS 1.3 Best Practice for Time and Elevation issued by OGC Meteorology/Oceanography Domain Working Group (https://portal.opengeospatial.org/files/?artifact_id=56394).
- If using commercial software that offers both proprietary services and open standards services, make sure to turn on options for standard services. For instance, the Esri ArcGIS Server platform has options for turning on OGC capabilities to provide WMS, WFS, and/or WCS services in addition to the Esri proprietary services, but these capabilities are not generally turned on by default.
- Large-file download remains an issue over the HTTP protocol and dependent on many networking factors. Where possible, employ bit streaming protocols such as BitTorrent or file subsetting capabilities to mitigate this issue.

Implementations

Software

The following software can be used to implement these services on your own data. *Note: inclusion on this list does not constitute formal endorsement by the U.S. Government; no warranty regarding suitability, security, or performance of software is expressed or implied.*

- DAP
    - THREDDS Data Server (http://www.unidata.ucar.edu/software/thredds/current/tds/) - free, open source.
    - Hyrax Data Server (http://www.opendap.org/download/hyrax) - free, open source.
    - ERDDAP (http://coastwatch.pfeg.noaa.gov/erddap/) - free, open source.
- WMS
    - MapServer (http://mapserver.org/) - free, open source.
    - ERDDAP (http://coastwatch.pfeg.noaa.gov/erddap/) - free, open source.
    - Geoserver (http://geoserver.org/) - free, open source.
    - Boundless OpenGeo Suite (http://boundlessgeo.com/solutions/opengeo-suite/) - not free; includes GeoServer and other components, built on open source.
    - Sci-WMS (Python WMS service for unstructured grid data distributed via OPeNDAP) (https://github.com/asascience-open/sci-wms) - open source.
    - ncWMS - Web Map Service (http://www.opengeospatial.org/standards/wms) for geospatial data that are stored in CF-compliant (http://www.cfconventions.org/) NetCDF (http://www.unidata.ucar.edu/software/netcdf/) files. ncWMS can be deployed as a stand-alone application or as a plugin

for a THREDDS Data Server (http://www.resc.rdg.ac.uk/trac/ncWMS/) - open source.

    - ESRI ArcGIS Server (if OGC option enabled) - not free.

- SOS
    - U.S. IOOS® customized build of the 52°North Sensor Observation Service (SOS Project page: http://ioos.github.io/i52n-sos/, Github Repository: https://github.com/ioos/i52n-sos) - free, open source.
    - ncSOS plugin for the THREDDS Data Server (https://github.com/asascience-open/ncSOS) - free, open source.
- TableDAP
    - ERDDAP (http://coastwatch.pfeg.noaa.gov/erddap/download/setup.html) - free, open source.

Instances

The following are examples of Federal use of these Data-access services.

- DAP
    - NOAA Operational Model Archive and Distribution System (NOMADS) - open source.
        - http://nomads.ncdc.noaa.gov/
    - NOAA Unified Access Framework (UAF) - open source.
        - https://geo-ide.noaa.gov/
    - USGS ScienceBase - dynamic query returning data assets with OPeNDAP end points.

        - https://www.sciencebase.gov/catalog/
- WMS
    - NOAA nowCoast - open source.
        - http://nowcoast.noaa.gov/help/mapservices.shtml?name=mapservices
    - Environmental Protection Agency (EPA) Environmental Dataset Gateway (EDG)
        - https://edg.epa.gov/metadata/catalog/main/home.page
    - USGS ScienceBase - dynamic query returning data assets with OPeNDAP end points.
        - https://www.sciencebase.gov/catalog/
- WFS
    - USGS ScienceBase - dynamic query returning data assets with OPeNDAP end points.
        - https://www.sciencebase.gov/catalog/
- WMTS
    - NASA Global Imagery Browse Services (GIBS) - open source.
        - https://earthdata.nasa.gov/about-eosdis/science-system-description/eosdis-components/global-imagery-browse-services-gibs

Testing

- OGC Compliance Program (http://cite.opengeospatial.org/) offers tools and services for testing implementations.
- The NSF Long-Term Ecological Research (LTER) Program has developed a method of testing “metadata congruence” (https://im.lternet.edu/projects/eml_congruency_checker)

that could be extended to test how the data served by a service matches the description in the metadata.

Metrics

- Percentage of metadata records or catalog entries with links to data-access services.
- Number of datasets that meet the U.S. Chief Information Officer qualifications for Open Data: publicly accessible under an open license with no restrictions on reuse. See the full Open Data Principles at https://project-open-data.cio.gov/principles
- Relative popularity among data providers of each type of data-access service.

Known Limitations

- Existing standards may need additional constraints, extensions, or profiles to provide more-consistent interoperability.
- Some types of biological data may not match any of the feature types listed above.
- Some types of social science/economics data may not match any of the feature types listed above.
- Different standards are needed for video or audio environmental observations (*e.g.*, for undersea video).

Data Documentation

Introduction

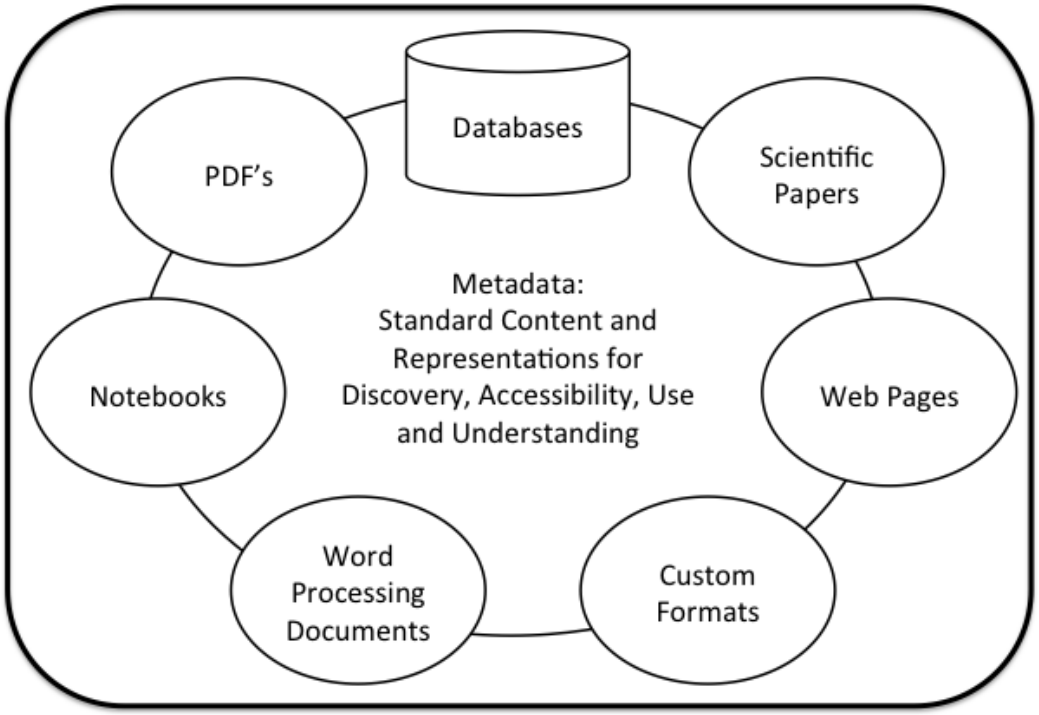
Many scientific datasets and products are documented using approaches and tools developed by scientists and data collectors to support their own analysis and understanding. Such documentation exists in laboratory notebooks, scientific papers, web pages, user guides, and other ad hoc formats, each with associated storage and preservation strategies. This customized and often unstructured approach may be sufficient for independent investigators or individuals working in the confines of a particular laboratory or community. However, it makes it difficult for users outside of these small groups to discover, access, use, and understand data without consulting the data creators.

Metadata (the collected information describing data's structure and source) provides well-defined content in structured representations. When complete and accurate, metadata allows users to access and quickly understand many aspects of datasets regardless of their source. Metadata can also be integrated into discovery and analysis tools, and can provide consistent reference to external documentation.

Metadata standards provide element names and associated structures that can describe a wide variety of digital resources. The definitions of these elements and accepted values are intended to be sufficiently broad to satisfy the metadata needs of various disciplines. These standards also include references to external documentation and well-defined mechanisms for adding structured information to address specific community needs.

The standards below, properly applied, can be used to meet the following Data Management Principle from the GEO Strategic Plan:

- **DMP-4**: Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international or community-approved standards. To the extent possible, data will also be described in peer-reviewed publications referenced in the metadata record.



Documentation: all information that supports effective sustainable utilization and reproducibility of scientific data and products

Standards and Protocols

There is growing convergence on a suite of international metadata standards and associated resources such as code lists and classification systems that are enabling interoperability among data systems and catalogs.

Metadata	
Model	Encoding Schema
ISO 19115-2	ISO 19139-2
ISO 19115	ISO 19139
ISO 19115-1 + 19157 (for quality)	ISO 19115-3

Note: FGDC CSDGM is accepted for legacy datasets that cannot comply with newly defined and adopted standards.

For instrument specifications:

- SensorML – SensorML provides a robust and semantically tied means of defining processes and processing components associated with the measurement and post-measurement transformation of observations.
    - http://www.opengeospatial.org/standards/sensorml

Methods and Practices

Support for Project Open Data metadata schema: for Earth Observations, preferred method is conversion from ISO XML (or other native format) to JSON (*e.g.*, via XML stylesheet - XSLT).

Implementations

- ncISO: A command-line utility for automating metadata analysis and ISO metadata generation for THREDDS (http://www.unidata.ucar.edu/software/thredds/current/tds/) Catalogs (http://www.ngdc.noaa.gov/eds/tds/) - open source.

Compatible Formats and Vocabularies

Introduction

A major goal of the Common Framework is to bring about improved interoperability among the data assets coming from observation systems. This approach requires the ability to write software and applications that combine, integrate, and synthesize observations. Interoperability requires compatibility among data systems at the format level, at least in terms of data-exchange formats. Interoperability also requires compatibility in data models and vocabularies so that, for instance, precipitation data from one source can be compared to rainfall data from another source. This aspect of the Common Framework focuses on the underlying compatibility among Earth-observing data systems, translation mechanisms, common data models, vocabularies, and semantics.

A fundamental decision point in developing the Common Framework was the determination that all data systems complying with this framework must support spatiotemporal queries and

constraints based on the Earth system. These queries could be either explicit, as in the case of an Earth-based coordinate reference system (or spatial reference system), or implicit, as in the case of a spatial framework such as hydrologic unit boundaries. For temporal constraints, parameters can be either explicit, based on standard encoding such as ISO8601 date/time variables, or implicit, such as geologic time periods. The space/time constraint or parameterization is assumed to be in place and available for discovering data, as in through catalog-service queries, and for accessing and using data, either through datasets being divided based on these parameters or through subsetting capabilities and available for use in various ways.

The standards below, applied properly, can be used to meet the following Data Management Principle from the GEO Strategic Plan:

- **DMP-3**: Data should be structured using encodings that are widely accepted in the target user community and aligned with organizational needs and observing methods, with preference given to non-proprietary international standards.

Standards and Protocols

Data Formats

- For numerical data:
    - NetCDF4 (Network Common Data Form, http://www.unidata.ucar.edu/software/netcdf/) – the most recent version developed by Unidata (http://www.unidata.ucar.edu/).
    - HDF5 (Hierarchical Data Format, http://www.hdfgroup.org) – the most recent version developed by the HDF Group.

- For imagery
    - GeoTIFF (http://trac.osgeo.org/geotiff/) – a public-domain metadata standard that allows georeferencing information to be embedded within a Tagged Image File Format (TIFF) file.
- For points/lines/polygons:
    - GML (Geography Markup Language, http://www.opengeospatial.org/standards/gml) – the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. Use of GML sometimes requires specific profiles in particular areas, especially various kinds of *in situ* measurements. For example:
        - For water-related data, the Common Framework recommends WaterML2.0 (http://www.opengeospatial.org/standards/waterml)
        - For weather data, the Common Framework recommends WXXM (Weather Information Exchange Model), http://www.wxxm.aero/public/subsite_homepage/homepage.html)

Controlled Vocabularies

Many communities of practice and research communities around specific domains have created controlled vocabularies to aid in ease of communication and proper reuse of data. Many of these vocabularies are too specific in applicability to broadly endorse, but we do strongly recommend the use of relevant controlled vocabularies as much as possible.

Spatial Reference System - There are many different spatial reference systems in use, and data developers sometimes create their own for various purposes. The recommendation of the Common Framework is to use:

- European Petroleum Survey Group (EPSG, http://www.epsg.org/) Geodetic Parameter Dataset (http://www.epsg-registry.org/).

Hydrologic Units – Hydrography is important for many uses of Earth-observation data and is complex enough to require its own reference system. The recommendation of the Common Framework is to use:

- National Hydrography Dataset (http://nhd.usgs.gov/),
    - Includes the Watershed Boundary Dataset available as a service (http://services.nationalmap.gov/arcgis/rest/services/nhd/MapServer) and for download (ftp://rockyftp.cr.usgs.gov/vdelivery/Datasets/Staged/WBD/).

Keywords – controlled lists of keywords for use in tagging data can aid in data search and discovery as well as interoperability. The recommendation of the Common Framework is to use:

- Office of Management and Budget Circular A-16 - Coordination of Geographic Information and Related Spatial Data Activities.
    - Reissued in August 2002.
    - Establishes FGDC with its structure and responsibilities.
    - The Appendices include definitions related to Geographic Information that are useful for universal adoption.
    - https://www.whitehouse.gov/omb/circulars_a016_rev/

- Global Change Master Directory (GCMD) – a lexicon developed by NASA of terms related to global-change research.
    - http://gcmd.nasa.gov/learn/keyword_list.html

Parameter Names –for naming data parameters in metadata fields, the recommendation of the Common Framework is to use:

- Climate Forecast Metadata Conventions Standard Names (http://cfconventions.org/standard-names.html)

Content Models

- U.S. Geoscience Information Network "community" content models (http://schemas.usgin.org/home/)
- Darwin Core (biological species observations, http://rs.tdwg.org/dwc/)
- NEPAnode (open geospatial web application for collaborating on data, maps, projects designed for non-GIS experts, http://nepanode.anl.gov/)

Methods and Practices

While having data in widely used formats is crucial, those formats can be used in ways that will not be compatible. It is important to test sufficiently in order to verify that content provided is readable by tools that users employ. The Common Framework also recommends validating file contents by using format and syntax checkers, if available.

Implementations

Building on the National Hydrography Dataset (NHD), USGS and EPA have collaborated on the production of the NHD Plus (http://www.horizon-systems.com/nhdplus/), which provides data down to the catchment level and is available via services from the EPA (http://water.epa.gov/scitech/datait/tools/waters/services/mapping_services.cfm) and download (http://www.horizon-systems.com/NHDPlus/V2NationalData.php) from Horizon Systems, a commercial partner in the NHD.

Appendix 1: Related Organizations and Policies

USGEO

The U.S. Group on Earth Observations (USGEO) is chartered as a Subcommittee of the White House National Science and Technology Council (NSTC) Committee on Environment, Natural Resources, and Sustainability (CENRS). USGEO is an interagency body, with thirteen government agencies represented. USGEO's purpose is threefold:

- to coordinate, plan, and assess Federal Earth observation activities in cooperation with domestic stakeholders;
- to foster improved Earth system data management and interoperability throughout the Federal Government; and,
- to engage interagency and national stakeholders to formulate the U.S. position for, and coordinate U.S. participation in the international Group on Earth Observations (GEO).

One activity of USGEO is carrying out the triennial Earth Observations Assessment (EOA), which assesses the impact of Earth observing systems as inputs to key products, services, and outcomes related to 13 Societal Benefit Areas. The EOA informs the *National Plan for Civil Earth Observations*. It has also informed agencies in identifying the high-impact datasets for which advancing data stewardship is a priority.

The USGEO Plan of Work for 2014-2016 includes as a priority action: "Establish criteria for assessing the relative capabilities in discoverability, accessibility, and usability of Earth observations data." This document was completed as the result of that action.

Data Management Working Group

The Data Management Working Group (DMWG) is chartered by and reports to USGEO. DMWG members, who are representatives of the various agencies within USGEO, developed this Common Framework document. The DMWG's charter calls for the group to foster “the implementation of an interagency framework for life cycle data management, stewardship and preservation for Earth-observation data.” The scope of work within its jurisdiction “includes data-management principles, approaches, data architecture considerations, data standardization, and data access and sharing policies...” The overall DMWG purpose is strengthening inter-agency access and sharing of Earth-observation data in order to extend and maximize agency investments made in data management and stewardship."

National Strategy for Civil Earth Observations

In April 2013, the National Science and Technology Council published the *National Strategy for Civil Earth Observations* (National Strategy), which promoted the importance of data

management and delivery. The National Strategy laid out a set of principles that all agencies should use when managing Earth-observation data. These principles are:

- **Full and open access**: Earth observations should be fully and openly available to all users promptly, in a nondiscriminatory and platform-agnostic manner, and generally free of charge wherever possible. Agencies may set user charges consistent with OMB Circular A-130.
- **Preservation**: Earth observations should be managed as an asset and preserved for future use.
- **Information data quality**: Earth observations should be of known quality and fully documented.
- **Ease of use**: Earth observations should be easily discoverable, searchable, and accessible online using interoperable services and standardized, machine-readable formats that encourage the broadest possible use.

The National Strategy also explained the benefit of using data-management standards:

> "Standards are common rules, conditions, guidelines, or characteristics for data and related processes, technology, and organization. The broad use of a small set of common data, metadata, and protocol standards across Federal agencies (and international standards where possible) for data development, documentation, and exchange enhances the utility of Earth observations, decreases the cost of using the data, and helps Federal agencies avoid redundancies and waste.

> “Different types of standards are applicable to various aspects of the data life cycle, including standards for data quality, metadata standards that specify the content and structure of documentation about a data set, and interoperability standards that specify the content and structure of the digital data and how services will interact. The use of existing national and international data, metadata, and protocol standards is recommended. Because such standards are often general-purpose and require specialization for specific data types, agencies are encouraged to publish the conventions, profiles, and examples they adopt to make these standards more applicable to their data. Collaboration among and within agencies is encouraged to adopt common conventions for using existing standards to maximize compatibility."

This document is the outcome of a collaboration of many agencies involved in USGEO to recommend common practices, and thus addresses this recommendation of the *National Strategy*. The complete *National Strategy for Civil Earth Observations* can be found here: https://www.whitehouse.gov/sites/default/files/microsites/ostp/nstc_2013_earthobsstrategy.pdf

National Plan for Civil Earth Observations

In July 2014, the Office of Science and Technology Policy published the *National Plan for Civil Earth Observations*. In the document were listed several supporting actions that were required to meet the Plan's priorities. The second-highest-priority supporting action was to “improve data access, management, and interoperability.” The Earth Observation Common Framework addresses many of the tasks that this action called upon agencies to undertake, including:

- Promote implementation of end-to-end life-cycle data management, including data-management principles, guidance, and data policy.

Encourage the development and use of uniform methodologies and practices across Federal agencies for common services in the handling of Earth-observation data to increase interoperability through improved metadata standardization. The complete *National Plan for Civil Earth Observations* can be found here: https://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/2014_national_plan_for_civil_earth_observations.pdf

The Big Earth Data Initiative

The Big Earth Data Initiative (BEDI) is a cross-agency activity supported by the Office of Science and Technology Policy (OSTP) and the Office of Management and Budget (OMB), and coordinated by the USGEO DMWG. Available BEDI funding is used to directly improve data from NASA, NOAA, and USGS, according to the recommendations of this Common Framework.

Guidance for BEDI comes from the April 2013 *National Strategy for Civil Earth Observations*, and is complementary to the May 2013 Executive Order "Making Open and Machine Readable the New Default for Government Information", and the associated Open Data Policy (OMB M-13-13). BEDI aims to improve the discovery, access, and use of all Federally held Earth-system data with a high impact for public- and private-sector decision-making. Its objectives are to: (1) maximize the availability of data and information and ensure dissemination in a timely and usable manner; (2) facilitate the transformation of observations and data into useful information through the use of open, machine-readable formats and Application Programming Interfaces

(APIs); (3) increase interoperability of Earth-system data and tools in order to encourage the development and use of tools and practices across Federal agencies; and (4) support the development of information products and tools that directly support decision-making.

Open Data Initiatives

OSTP Requirement for Agency Public Access Plans

In February 2013, White House Office of Science and Technology Policy (OSTP) Director John Holdren directed Federal agencies with more than $100M per year in R&D expenditures to develop plans for increasing access to the results of federally funded research. Plans are to ensure that peer-reviewed scholarly publications are freely available to the public within one year of publication. Agency plans must also ensure that researchers better account for and manage the digital data resulting from federally funded scientific research. Among the objectives for public access to scientific data are that agencies' plans should:

- Maximize access, by the general public and without charge, to digitally formatted scientific data created with Federal funds.
- Ensure the development and appropriate evaluation of data management plans describing proposed procedures for long-term preservation of, and access to, scientific data.
- Include mechanisms to ensure that intramural and extramural researchers comply with data management plans and policies.
- Promote the deposit of data in publicly accessible databases, where appropriate and available.

- Develop approaches for identifying and providing appropriate attribution to scientific data sets that are made available under each agency's plan.

The full public-access memo can be found here:

http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Office of Management and Budget Open Data Policy

On May 9, 2013, the Office of Management and Budget (OMB) issued memorandum M-13-13 with the subject "Open Data Policy – Managing Information as an Asset." This memorandum said, in part:

> "Specifically, this Memorandum requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities. This includes using machine-readable and open formats, data standards, and common core and extensible metadata for all new information creation and collection efforts. It also includes agencies ensuring information stewardship through the use of open licenses and review of information for privacy, confidentiality, security, or other restrictions to release. Additionally, it involves agencies building or modernizing information systems in a way that maximizes interoperability and information accessibility, maintains internal and external data asset inventories, enhances information safeguards, and clarifies information management responsibilities."

The full memorandum can be found

here: http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

Group on Earth Observations Data Management Principles

In 2014, the international Group on Earth Observations (GEO) Data Management Principles Task Force (DMP-TF) developed a set of principles intended to maximize the value and benefit of data shared through GEOSS. The following principles were adopted by GEO in 2015 as part of its Strategic Plan. The USGEO Common Framework directly supports Data Management Principles #1, 2, 3, 4, and 10:

- **Discoverability**
    - DMP-1: Data and all associated metadata will be discoverable through catalogs and search engines, and data access and use conditions, including licenses, will be clearly indicated.
- **Accessibility**
    - DMP-2: Data will be accessible via online services, including, at minimum, direct download but preferably user-customizable services for visualization and computation.
- **Usability**
    - DMP-3: Data should be structured using encodings that are widely accepted in the target user community and aligned with organizational needs and observing methods, with preference given to non-proprietary international standards.
    - DMP-4: Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured

metadata based on international or community-approved standards. To the extent possible, data will also be described in peer-reviewed publications referenced in the metadata record.

  - DMP-5: Data will include provenance metadata indicating the origin and processing history of raw observations and derived products, to ensure full traceability of the product chain.
  - DMP-6: Data will be quality-controlled and the results of quality control shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.

- **Preservation**
  - DMP-7: Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.
  - DMP-8: Data and associated metadata held in data management systems will be periodically verified to ensure integrity, authenticity and readability.
- **Curation**
  - DMP-9: Data will be managed to perform corrections and updates in accordance with reviews, and to enable reprocessing as appropriate; where applicable this shall follow established and agreed procedures.
  - DMP-10: Data will be assigned appropriate persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement of use of their data.

As a member of GEO, the U.S. Government supports the goal of full implementation of these principles.

Definitions and Descriptions

API (Application Programming Interface) – a set of routines, protocols, and tools for building software applications, which expresses a software component in terms of its operations, inputs, outputs, and underlying types to allow automated access to machine-readable information.

BEDI (Big Earth Data Initiative) – a cross-agency budget initiative aiming to improve Earth-observation data discoverability, accessibility, and usability.

Data – in this document, refers to Earth-observation data, meaning geo-referenced digital information about Earth, including the observations, metadata, imagery, derived products, data-processing algorithms (including computer source code and its documentation), and forecasts and analyses produced by computer models. Non-digital data, published papers, preserved geological or biological samples, or other media that have not been digitized are not included in this definition.

CENRS Committee on Environment, Natural Resources, and Sustainability (part of the National Science and Technology Council).

CFEOD (Common Framework for Earth-Observation Data) – this document.

DAP (Data Access Protocol) – a data-transmission protocol designed specifically for science data relying on HTTP and MIME standards.

DMP (Data Management Principle) – principles adopted by GEOSS to guide data stewardship.

DMWG (Data Management Working Group) – a working group under the USGEO Subcommittee of the CENRS.

FGDC (Federal Geographic Data Committee) – an interagency body within the U.S. Government.

GEO (Group on Earth Observations) – an international intergovernmental body that coordinates Earth-observation and data systems.

GEOSS (Global Earth Observation System of Systems) – a system being developed by GEO to deliver Earth-observation data to end users and decision makers.

HDF (Hierarchical Data Format) – a suite of open-source technologies to support scientific data.

ISO (International Organization for Standardization) – an independent, non-governmental membership organization that develops voluntary standards.

NetCDF (Network Common Data Form) – a set of software libraries and self-describing, machine-independent data formats.

NSTC - National Science and Technology Council.

OGC (Open Geospatial Consortium) – an international industry consortium of companies, government agencies, and universities developing publicly available interface standards.

USGEO (U.S. Group on Earth Observations) – an interagency subcommittee of the CENRS.

XML (Extensible Markup Language) – a language that defines a set of rules for encoding documents in a format that is both human- and machine-readable.